An Extension of Shapiro-Wilk's Test for Multivariate Normality and Power Investigation for Contaminated Alternatives

R. Sakthivel[a] and Martin L. William[b]

[a]Department of Statistics, Presidency College, Chennai, India
[b]Department of Statistics, Loyola College, Chennai, India

Abstract- Many procedures are available in the literature on testing for univariate normality. Among them, the procedure of Shapiro and Wilk (1965) is a very effective test to detect departures from normality. The Shapiro-Wilk procedure has been extended to test multivariate normality recently by Alva and Estrada (2009). The present paper proposes another approach to extend the Shapiro-Wilk procedure for testing multivariate normality. A simulation study has been carried out to get the critical values of the proposed statistic in order to make it applicable to dimensions 2 and 5. Power comparison of the new approach to the one given by Alva and Estrada (2009) is presented for a contaminated alternatives. The software for the same is developed in R Language.

Index Terms- Testing Multivariate normality, Power of Tests, Monte Carlo Simulation

I. Introduction

In most of the applications of statistics dealing with multi-dimensional datasets, the multivariate normal distribution plays a pivotal role. Much theoretical development has taken place so far on the multivariate normal distribution and many of the procedures for analyzing multivariate response data require joint normality of the multivariate responses. Simulation studies of Mardia (1975), besides those of others, emphasize the importance of the multivariate normality assumption for many of these procedures which are not robust to non-multivariate-normal data.

In the past, about 40 to 50 different approaches have been proposed to test for multivariate normality, but no single procedure is the best to detect various types of departures from multivariate normality. An early work is that of Healy (1968) who proposed a plotting technique while Andrews *et al.* (1971) gave a procedure based on Mahalanobis distances. A powerful procedure based on the Shapiro-Wilk (1965) procedure was given by Royston (1983). Recent contributions in this area include the articles by Szekely and Rizzo (2005), Holgersson (2006), Alva and Estrada (2009) and Tenreiro (2011) among others. Even though many tests for multivariate normality exist, the assumption of multivariate normality is generally not tested by many practitioners of statistics. This is due to lack of awareness and non-availability of software for many of these techniques.

In this paper, we seek to extend the Shapiro-Wilk procedure because this test is very powerful compared to many other procedures in the univariate case. Alva and Estrada (2009) have recently proposed one extension of the Shapiro-Wilk procedure to test for multivariate normality. These authors, through numerical simulation studies, investigated the power of their test against a wide range of alternatives for dimensions p = 2 and 5. Although their test is better than other existing procedures in detecting some types of departures from multivariate normality, it does not work well in detecting contamination from non-normal distributions especially when p = 5.

In this paper, we propose a different approach to extend the Shapiro-Wilk procedure to test for multivariate normality. For now, we consider dimensions 2 and 5 and aim at detecting certain types of non-normal contaminations in multivariate distributions. As there are no theoretical / algebraic expressions for the distribution of the Shapiro-Wilk statistic, the present study uses Monte Carlo simulations for developing tables of critical values and for carrying out power investigations as Alva and Estrada (2009) have done.

In Section 2, the approach of Alva and Estrada (2009) is briefly discussed. The proposed new extension of Shapiro-Wilk's approach is presented in Section 3. The critical values of the test of size $\alpha = 0.05$ for dimensions 2 and 5 are developed in Section 4. Section 5 discusses the results of the power investigations carried out for a class of contaminated alternatives and compares the powers of the proposed test procedure with the test of Alva and Estrada (2009). In Section 6, remarks highlighting the findings of the current investigations are given. The software for the new approach has been developed using R Language.

II. A Review of the Procedure of Alva-Estrada

Let $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_n}$ be independent and identically distributed random vectors of dimension p ≥ 1. Denote the p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and var-cov matrix $\boldsymbol{\Sigma}$ by $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Also, let $\mathbf{0}$ be the null vector of order p and $\mathbf{I}$ be the identity matrix of order p x p. It is well known that $\mathbf{X_i} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if and only if $\mathbf{Z_i} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X_i} - \boldsymbol{\mu}) \sim N_p(\mathbf{0}, \mathbf{I})$.

Let $\overline{\mathbf{X}} = n^{-1}\sum_{i=1}^{n}\mathbf{X}_i$ be the sample mean vector and $\mathbf{S} = n^{-1}\sum_{i=1}^{n}(\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})'$ be the sample var-cov matrix and $\mathbf{S}^{-1/2}$ be the positive definite square root of $\mathbf{S}^{-1}$. When $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_n}$ follow $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the random vectors $\mathbf{Z_i} = \mathbf{S}^{-1/2}(\mathbf{X_i} - \overline{\mathbf{X}})$ have a distribution close to $N_p(\mathbf{0}, \mathbf{I})$. This also means that the coordinates of $\mathbf{Z_i}$ denoted $Z_{1i}, Z_{2i}, \ldots, Z_{pi}$ are approximately independent univariate normal random variables.

To test the null hypothesis H_0: $\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_n}$ is a sample from p-variate normal distribution, Alva and Estrada (2009) proposed the test statistic

$$W_{AVE} = p^{-1} \sum_{i=1}^{p} W_{Zi} \quad (2)$$

where W_{Zi} is the Shapiro-Wilk's statistic evaluated on the i^{th} coordinate of the new coordinate system $\mathbf{Z_i}$. If H_0 were true, each of the W_{Zi} is expected to be close to unity. Alva and Estrada (2009) therefore, proceed with the understanding that if W_{AVE} is close to unity, each W_{Zi} is expected to be close to unity. Thus, the test would reject H_0 for 'small' values of the statistic W_{AVE}. That is, the test rejects H_0 if $W_{AVE} < c_{\alpha;n,p}$ where $c_{\alpha;n,p}$ is a quantity satisfying the condition

$$P\{W_{AVE} < c_{\alpha;n,p} \mid H_0\} = \alpha$$

α being the desired size of the test. It is noted that for $p = 1$, the Alva-Estrada statistic reduces to the Shapiro-Wilk statistic.

III. THE PROPOSED EXTENSION OF SHAPIRO-WILK PROCEDURE

The present paper develops a modified extension of Shapiro-Wilk procedure, different from that of Alva and Estrada (2009). The procedure suggested by these authors is based on the 'Average' of the Shapiro-Wilk statistics computed in the 'p' dimensions. However, the averaging of the 'p' statistics may probably 'hide' the presence of a 'small' value among them and thereby lead us to wrongly accept joint normality when it is actually not the case. Thus, the existence of a non-normal marginal in one or a few of the dimensions may go undetected in the approach of Alva-Estrada. The proposed modification is expected to be more sensitive in detecting marginal departures from normality.

The statistic we propose in this paper is

$$W_{MIN} = \underset{1 \le i \le p}{Min} \left\{ W_{Zi} \right\} \quad (3)$$

If H_0 were true, 'each' of the W_{Zi} is expected to be close to unity; equivalently, the least of the W_{Zi} is expected so. Thus, the test would reject H_0 for 'small' values of the statistic W_{MIN}. That is, the test rejects H_0 if $W_{MIN} < d_{\alpha;n,p}$ where $d_{\alpha;n,p}$ is a quantity satisfying the condition

$$P\{W_{MIN} < d_{\alpha;n,p} \mid H_0\} = \alpha$$

α being the desired size of the test. It is noted that for $p = 1$, the proposed new statistic reduces to the Shapiro-Wilk statistic.

IV. CRITICAL VALUES FOR THE WMIN TEST STATISTIC

As had been done for the earlier statistics mentioned, the null distribution of the statistic W_{MIN} also has to be obtained by simulation to find the critical values required for carrying out the test for multivariate normality. For dimensions p = 2 and 5 and for each of various sample sizes, we repeat the sampling 100000 times to generate the null distributions of the statistic W_{MIN} for each combination of 'p' and 'n'. Tracking the important quantiles identified over the various stages of repetitions, such as 1000, 5000, 10000, 50000, 75000 and 100000, it is found that 'convergence' take place in a convincing manner as the number of repetitions increase. In fact, the 'convergence' happens somewhere close to 10000 repetitions itself.

Even though, the quantiles of various orders found from the simulation study can be presented here, it is the quantiles of order 0.05 namely $d_{0.05;n,p}$ that are required for purposes of carrying out the test for multivariate normality and hence, quantiles of this order alone are summarized in Table 1 below. We emphasize that, the sample size 'n' is required to be greater than the dimension 'p' to carry out the test and therefore, the cells corresponding to n = 3, 4, 5 are empty when p = 5.

Table 1 [Critical values $d_{0.05;n,p}$ of W_{MIN}]

N	W_{MIN}(p=2)	W_{AVE}(p=2)	W_{MIN}(p=5)	W_{AVE}(p=5)
3	0.7609	0.8159	--	--
4	0.727	0.8077	--	--
5	0.7434	0.8169	--	--
6	0.7592	0.8285	0.7203	0.8598
7	0.7783	0.8411	0.7387	0.869
8	0.7943	0.8521	0.7555	0.8773
9	0.8071	0.8611	0.7721	0.8851
10	0.8188	0.8696	0.7858	0.8918

11	0.8304	0.8774	0.7979	0.898
12	0.8403	0.8843	0.8095	0.9036
13	0.8476	0.8899	0.8189	0.9084
14	0.8559	0.8956	0.8286	0.9127
15	0.8626	0.9003	0.8366	0.9167
16	0.8685	0.9046	0.8439	0.9204
17	0.8748	0.9092	0.8503	0.9237
18	0.8799	0.9123	0.857	0.9266
19	0.8847	0.9161	0.8617	0.9291
20	0.8891	0.9194	0.8673	0.932
30	0.9189	0.9406	0.9035	0.9499
40	0.9359	0.953	0.9239	0.9601
50	0.9468	0.9608	0.9372	0.9668
60	0.9545	0.9663	0.9461	0.9715
70	0.96	0.9704	0.9528	0.9749
80	0.9644	0.9736	0.9579	0.9776
90	0.9679	0.9761	0.9621	0.9798
100	0.9709	0.9783	0.9656	0.9815

V. POWER COMPUTATION AND COMPARISON OF W_{MIN} AND W_{AVE}

The powers of W_{MIN} and W_{AVE} have been estimated by using Monte Carlo simulation of samples of sizes n = 3 (1) 20 (10) 100 for dimension p = 2 and n = 6 (1) 20 (10) 100 for dimension p = 5. The test size is taken to be $\alpha = 0.05$. For each combination of (n, p), 10000 random samples were simulated from each of the specified alternative distribution. The alternatives that have been considered in this paper include multivariate distributions which are normal in some dimensions but 'contaminated' with non-normal distributions in one dimension. The findings are presented in the sequel.

5.1 Power investigations for dimension p = 2

The alternative distributions considered in this section are those with a normal marginal in one dimension and a non-normal marginal in the second dimension, as specified below:

(1) $N_1(0,1)$ x $Cauchy_1(0,1)$
(2) $N_1(0,1)$ x $Exponential_1(1)$
(3) $N_1(0,1)$ x $LogNormal_1(0,1)$
(4) $N_1(0,1)$ x $Weibull_1(1,1)$

The empirical powers of the procedures based on the proposed W_{MIN} statistic and the existing W_{AVE} statistic of Alva and Estrada (2009) are both reported for the different sample sizes and for the four different alternatives in the following table.

Table 2 [Empirical Powers of W_{MIN} and W_{AVE} for p = 2 and test size $\alpha = 0.05$]

n	WS_{MIN} (1)	WS_{AVE} (1)	WS_{MIN} (2)	WS_{AVE} (2)	WS_{MIN} (3)	WS_{AVE} (3)	WS_{MIN} (4)	WS_{AVE} (4)
3	0.0643	0	0.0488	0	0.0472	0	0.0505	0
4	0.1575	0.1098	0.0679	0.0538	0.0912	0.0697	0.0696	0.0619
5	0.2468	0.2115	0.097	0.09	0.1542	0.1313	0.0994	0.0863
6	0.3083	0.2921	0.1266	0.1216	0.2201	0.2074	0.1307	0.131
7	0.3814	0.3616	0.1683	0.1627	0.292	0.2744	0.1764	0.1642
8	0.4448	0.4261	0.2232	0.2128	0.3624	0.3467	0.2167	0.2093
9	0.5014	0.4852	0.2628	0.252	0.4272	0.3989	0.2584	0.2452

10	0.5359	0.5237	0.3112	0.298	0.4849	0.4686	0.3006	0.289
11	0.5874	0.57	0.3589	0.334	0.5431	0.5183	0.3536	0.3336
12	0.6319	0.6178	0.3982	0.3755	0.6038	0.5743	0.403	0.3793
13	0.6596	0.6432	0.4407	0.4157	0.6529	0.6298	0.4326	0.4123
14	0.6944	0.6824	0.4837	0.4597	0.7019	0.6753	0.492	0.4615
15	0.7237	0.7082	0.5308	0.4973	0.7386	0.7123	0.525	0.5014
16	0.7534	0.7426	0.5637	0.5357	0.7712	0.7443	0.5673	0.5374
17	0.7749	0.7672	0.6107	0.578	0.8029	0.7738	0.6103	0.5811
18	0.8033	0.7921	0.6502	0.6107	0.8318	0.8088	0.6474	0.6076
19	0.8222	0.8097	0.6892	0.6568	0.8565	0.8349	0.6783	0.6449
20	0.8421	0.8313	0.7169	0.6849	0.8846	0.8655	0.7165	0.6874
30	0.9455	0.9402	0.9161	0.8898	0.9803	0.9727	0.9146	0.8878
40	0.98	0.9782	0.9827	0.9725	0.9981	0.9965	0.9807	0.9715
50	0.9943	0.9938	0.9976	0.9951	0.9999	0.9995	0.9966	0.9954
60	0.9985	0.998	0.9998	0.9992	1	1	0.9993	0.9989
70	0.9999	0.9996	0.9999	0.9998	1	1	1	1
80	0.9999	0.9998	1	1	1	1	1	0.9999
90	0.9999	0.9999	1	1	1	1	1	1
100	1	1	1	1	1	1	1	1

The following figures give the power curves as a function of the sample size:

Fig 1 [Power Curves for $N_1(0,1)$ x Cauchy (0,1) Alternative based on 10000 Repetitions]

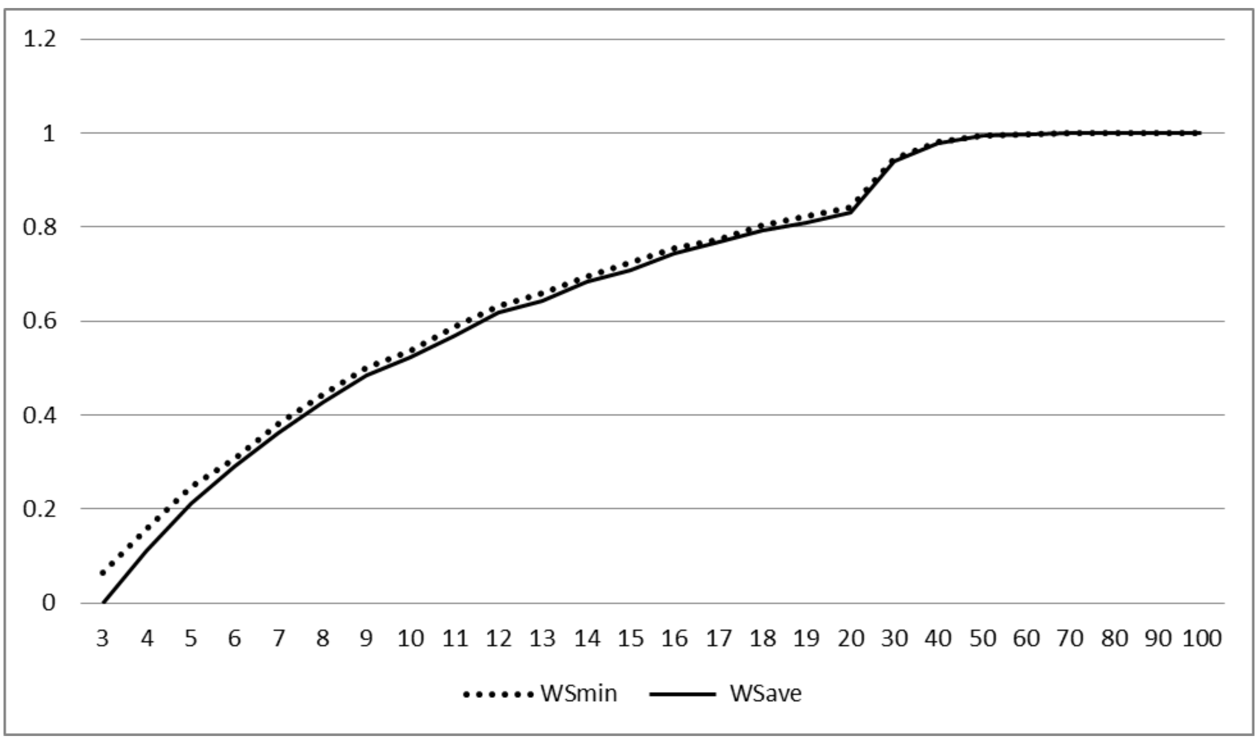


Fig 2 [Power Curves for $N_1(0,1)$ x $Exponential_1(1)$ Alternative based on 10000 Repetitions]

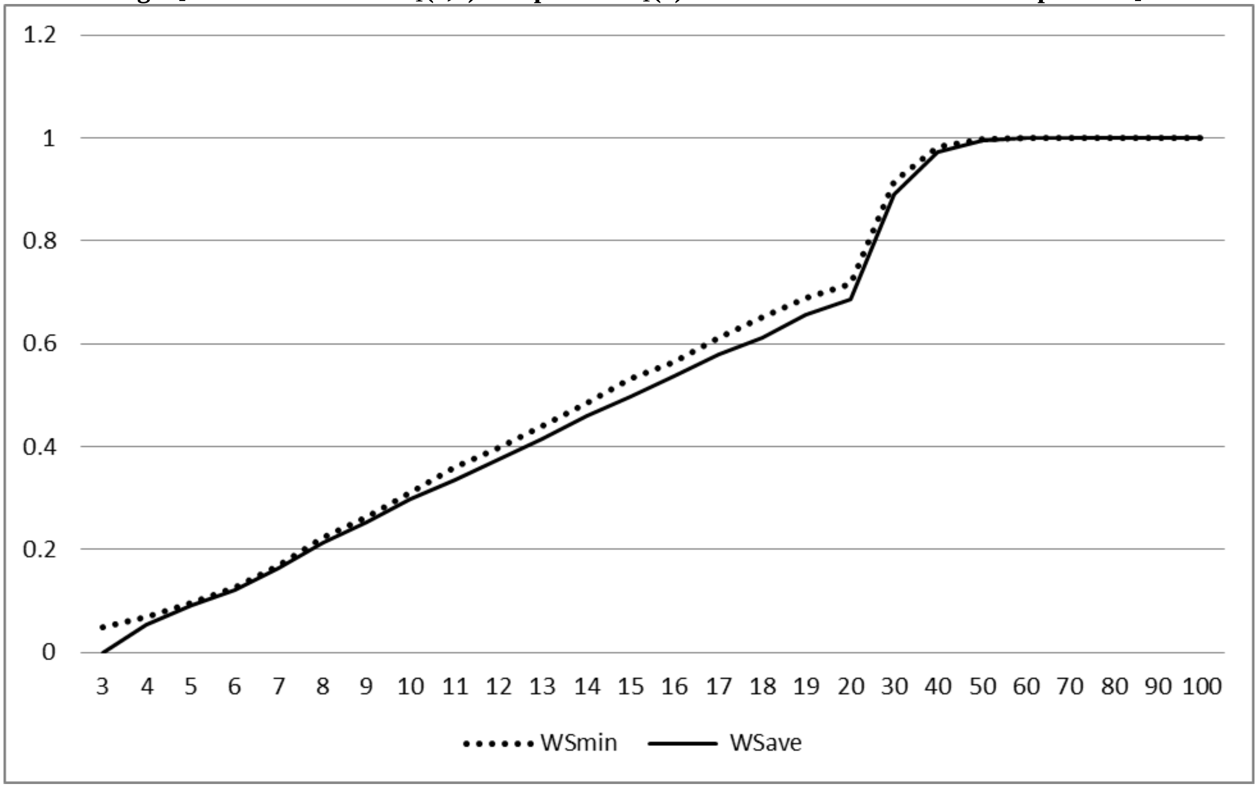


Fig 3 [Power Curves for $N_1(0,1)$ x $LogNormal_1(0,1)$ Alternative based on 10000 Repetitions]

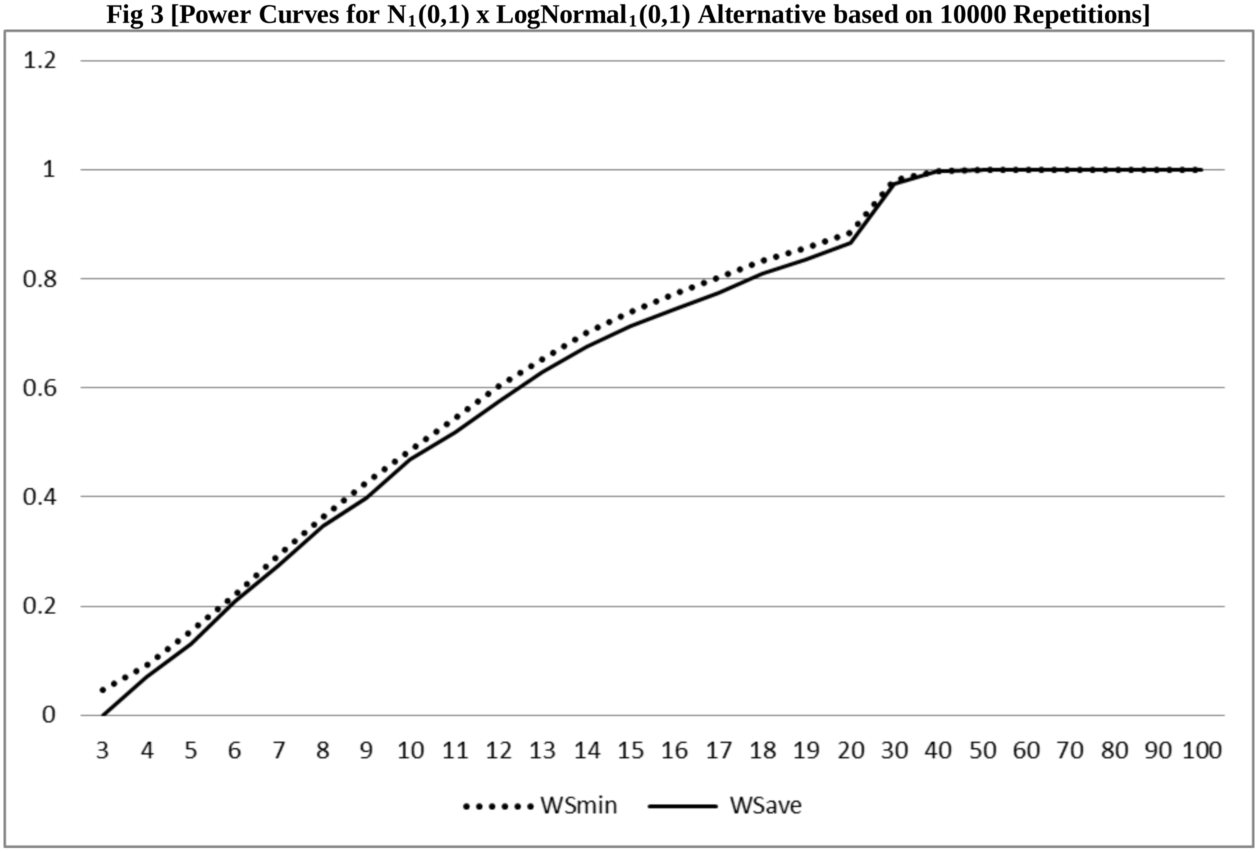
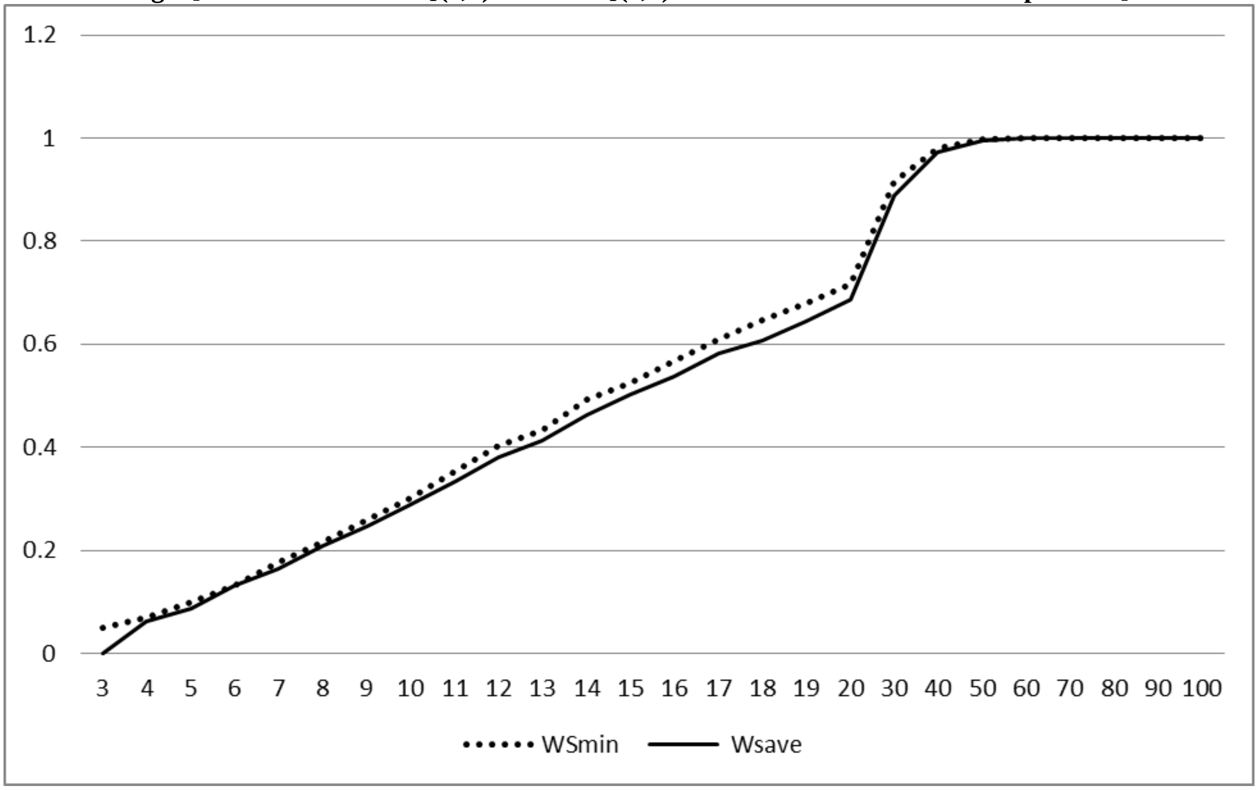


Fig 4 [Power Curves for $N_1(0,1)$ x $Weibull_1(1,1)$ Alternative based on 10000 Repetitions]



5.2 Power investigations for dimension p = 5

The alternative distributions considered in this section are those with normal marginals in four dimensions and a non-normal marginal in the fifth dimension, as specified below:

(5) $N_4(0,1)$ x $Cauchy_1(0,1)$
(6) $N_4(0,1)$ x $Exponential_1(1)$
(7) $N_4(0,1)$ x $LogNormal_1(0,1)$
(8) $N_4(0,1)$ x $Weibull_1(1,1)$

The empirical powers of the procedures based on the proposed W_{MIN} statistic and the existing W_{AVE} statistic of Alva and Estrada (2009) are both reported for different sample sizes and for the four different alternatives in the following table.

Table 3 [Empirical Powers of W_{MIN} and W_{AVE} for p = 5 and test size α = 0.05]

N	WS_{MIN} (5)	WS_{AVE} (5)	WS_{MIN}(6)	WS_{AVE}(6)	WS_{MIN} (7)	WS_{AVE} (7)	WS_{MIN} (8)	WS_{AVE} (8)
6	0.2544	0.1765	0.0664	0.0516	0.1191	0.081	0.0639	0.0488
7	0.3096	0.241	0.0786	0.0687	0.1537	0.1209	0.0828	0.0699
8	0.3681	0.3117	0.099	0.0879	0.2138	0.1754	0.0928	0.0843
9	0.4157	0.369	0.1273	0.1126	0.2597	0.2178	0.1254	0.1126
10	0.4806	0.4227	0.1516	0.135	0.3181	0.2681	0.153	0.1322
11	0.5165	0.4681	0.1835	0.1577	0.3693	0.3172	0.1744	0.1554
12	0.5545	0.5115	0.2123	0.1871	0.4189	0.3558	0.2152	0.186
13	0.6026	0.5495	0.2387	0.2128	0.4689	0.4026	0.2448	0.2174
14	0.6378	0.5931	0.2789	0.2332	0.5183	0.445	0.276	0.2327
15	0.6728	0.6222	0.3036	0.2587	0.5742	0.501	0.3079	0.2612
16	0.6934	0.6497	0.3562	0.3005	0.6185	0.542	0.3474	0.2942

17	0.7187	0.6756	0.3811	0.3234	0.6585	0.5756	0.3819	0.322
18	0.7429	0.6956	0.4189	0.3559	0.6921	0.6022	0.4212	0.3583
19	0.7696	0.7298	0.448	0.3718	0.7233	0.6379	0.4446	0.3701
20	0.7934	0.7552	0.4833	0.4109	0.7481	0.6698	0.4941	0.412
30	0.9278	0.9048	0.7825	0.6795	0.9429	0.8929	0.7712	0.6769
40	0.974	0.964	0.9232	0.8478	0.9903	0.9732	0.93	0.8489
50	0.9929	0.9871	0.9805	0.9419	0.9988	0.9933	0.9806	0.9401
60	0.9982	0.9957	0.9955	0.9789	0.9995	0.9984	0.9963	0.9797
70	0.9993	0.9986	0.9993	0.994	1	0.9999	0.9992	0.9938
80	0.9999	0.9998	0.9998	0.9988	1	0.9999	0.9999	0.9984
90	1	1	1	0.9999	1	1	1	0.9998
100	1	1	1	1	1	1	1	1

The following figures give the power curves as a function of the sample size:

Fig 5 [Power Curves for $N_4(0,1)$ x Cauchy (0,1) Alternative based on 10000 Repetitions]

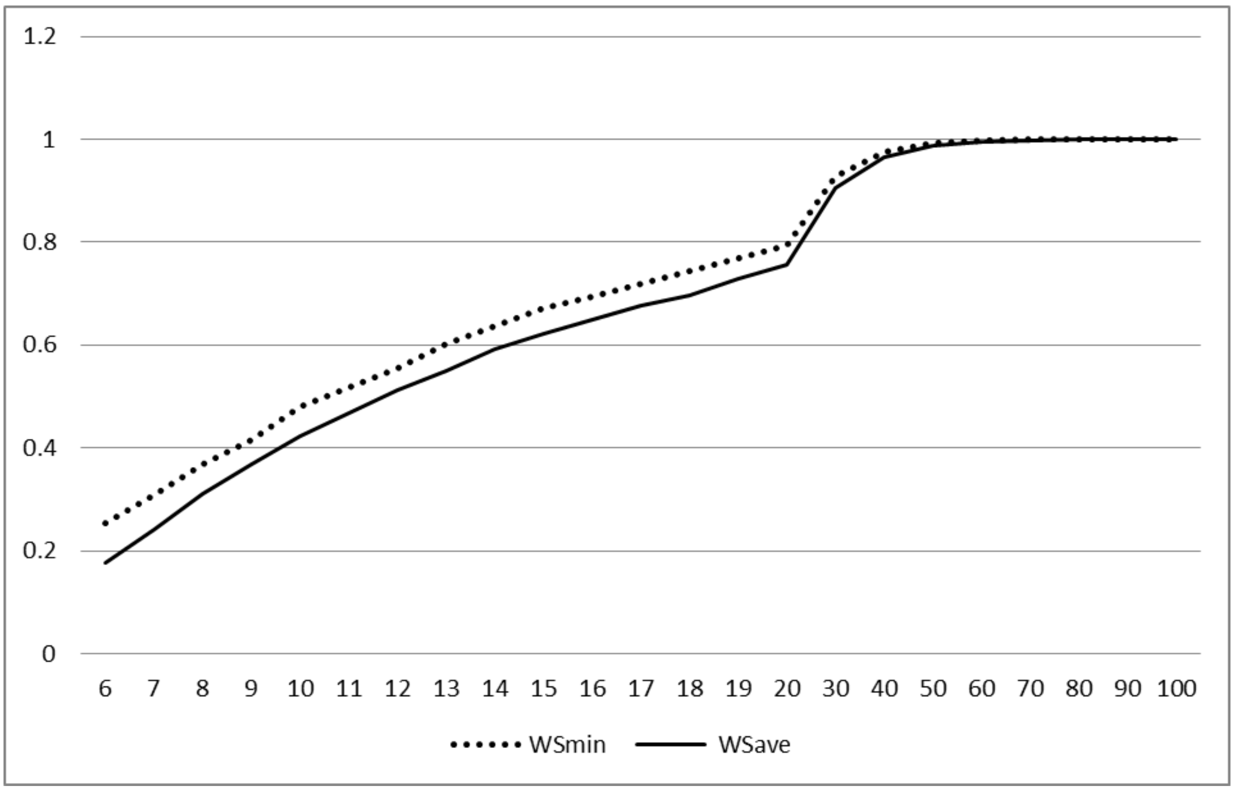


Fig 6 [Power Curves for $N_4(0,1)$ x $Exponential_1(1)$ Alternative based on 10000 Repetitions]

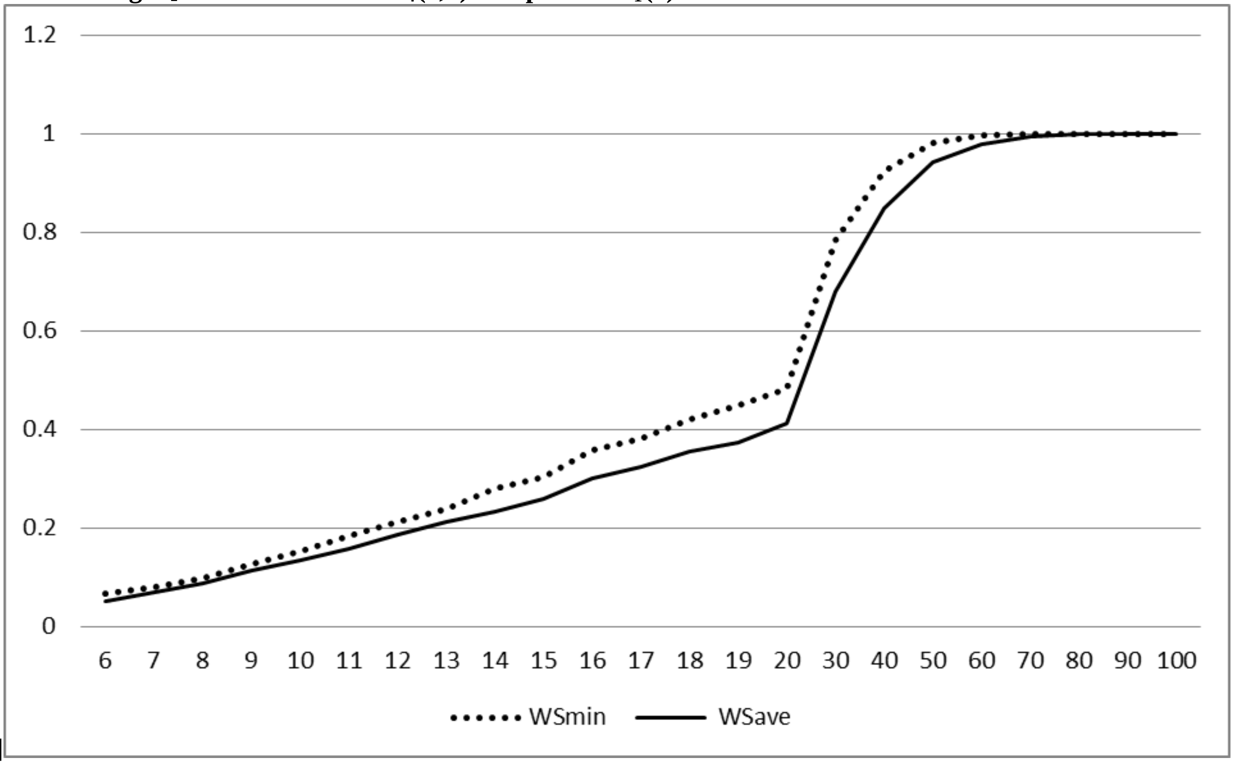


Fig 7 [Power Curves for $N_4(0,1)$ x $LogNormal_1(0,1)$ Alternative based on 10000 Repetitions]

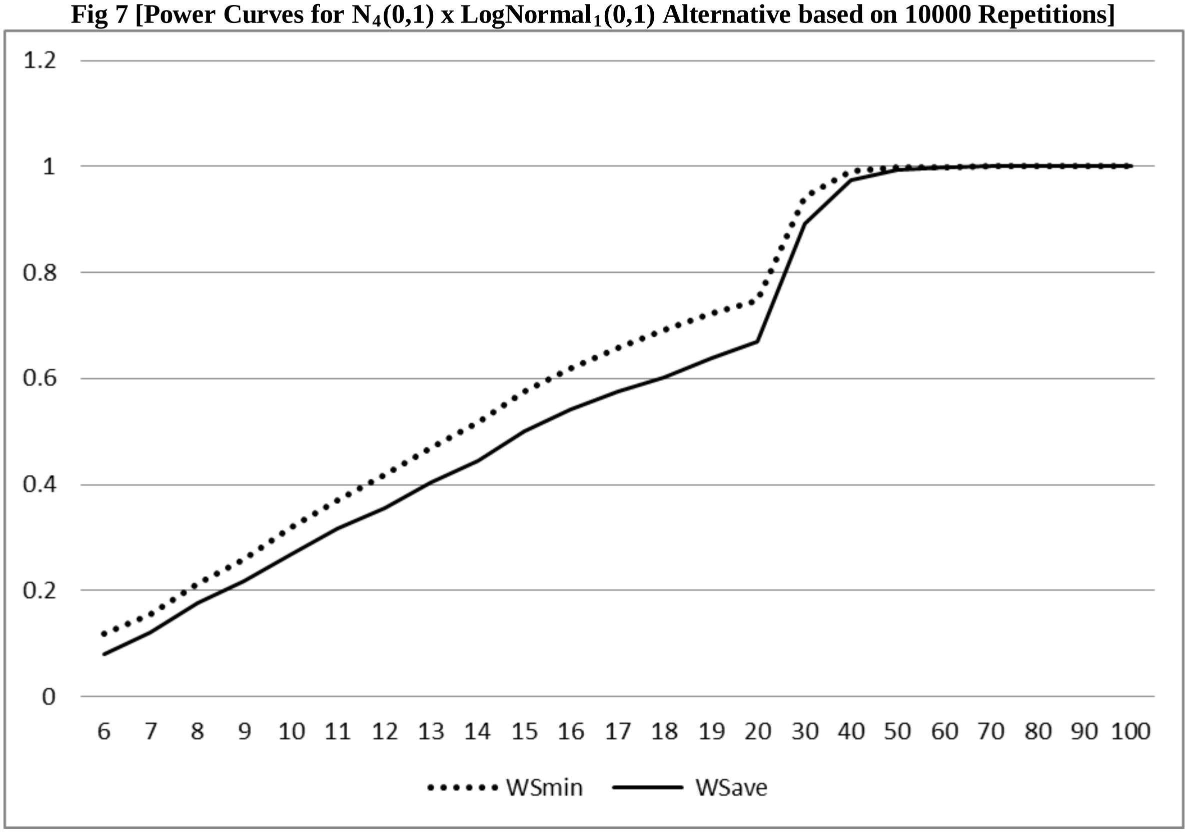
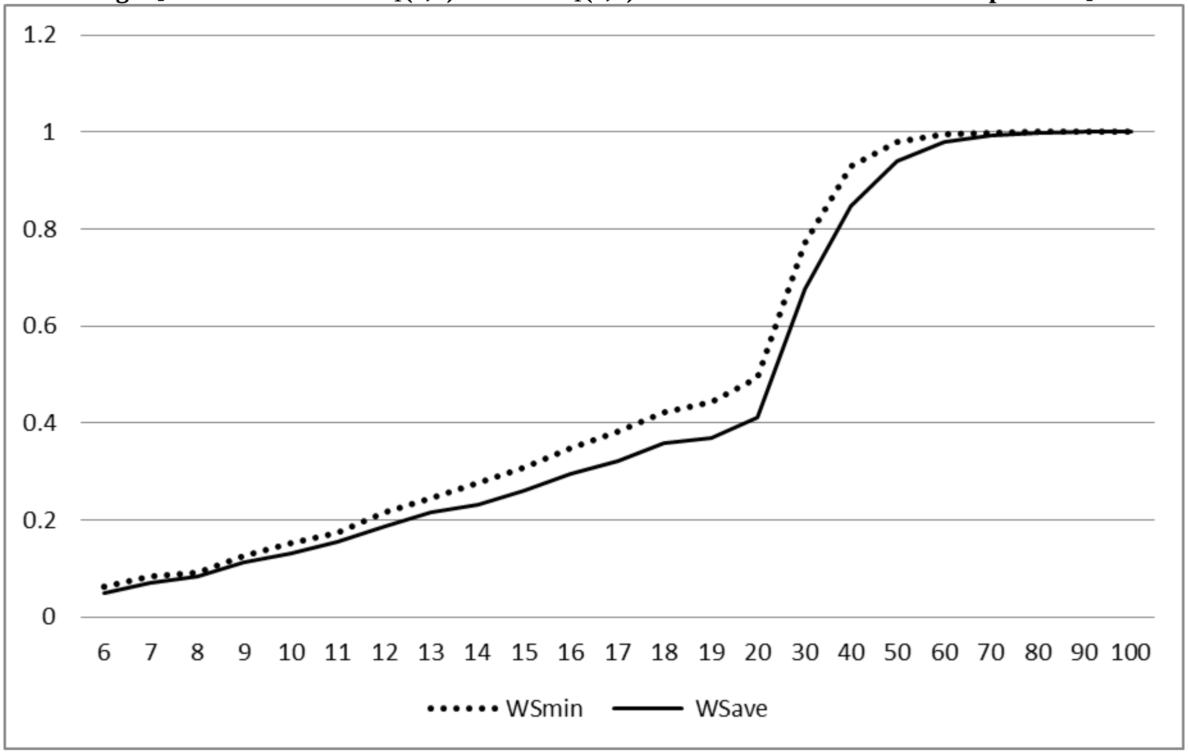


Fig 8 [Power Curves for $N_1(0,1)$ x $Weibull_1(1,1)$ Alternative based on 10000 Repetitions]



VI. CONCLUSIONS

A perusal of the findings in Section 5 brings out the fact that the new procedure based on W_{MIN} statistic is more sensitive in detecting even 'marginal' departures from normality compared to the one based on W_{AVE}. For bivariate distributions (p = 2), it is observed that the W_{MIN} procedure consistently performs better than the W_{AVE} procedure across all 'small sample' sizes while 'closeness' between the two procedures occurs for large sample sizes 'n' exceeding 40. Interestingly, for very small sample sizes, especially in the case of contamination from Cauchy and LogNormal marginals, the new procedure performs quite remarkably.

In the case of five-dimensional distributions, the consistently higher performance of the W_{MIN} procedure is more pronounced than for bivariate disributions. Also, as in the case of bivariate distributions, the remarkable higher performance of W_{MIN} for very small sample sizes is again observed in the case of Cauchy and LogNormal contaminations. The closeness between the W_{MIN} and W_{AVE} procedures occurs only for still larger sample sizes 'n' exceeding 60.

Overall, it is found that, the W_{MIN} procedure looks a promising one in detecting departures from normality in the form of non-normal marginal contaminations especially with samples of small and medium sizes. The performance of this statistic in detecting other types of non-normality in multivariate distributions is being addressed in another publication.

REFERENCES

[1] Alva Jose A. V. and Estrada E. G. (2009). A Generalization of Shapiro-Wilk's Test for Multivariate Normality. Communications in Statistics, 38: 1870 - 1883

[2] Andrews, D. F., Gnanadesikan, R and Warner, J. L. (1971). Transformations of multivariate data. Biometrics, 27: 825 - 840

[3] Healy, M. J. R. (1968). Multivariate normal plotting. Applied Statistics, 17: 157 - 161

[4] Holgersson, H. A graphical method for assessing multivariate normality. Computational Statistics, 21: 141 - 149

[5] Mardia, K. V. (1975). Assessment of multinormality and the robustness of Hotelling's T2. Applied Statistics, 24: 163 - 171

[6] Royston, J. P. (1983). Some techniques for assessing multivariate normality based on the Shapiro-Wilk W. Applied Statistics, 32: 121 - 133

[7] Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality. Biometrika, 52: 591 - 611

[8] Szekely, G. J. and Rizzo, M. L. (2005). A new test for multivariate normality. Journal of Multivariate Analysis, 93: 58 - 80

[9] Tenreiro, C. (2011). An affine invariant multiple test procedure for assessing multivariate normality. Computational Statistics and Data Analysis, 55: 1980 - 1992

AUTHORS

First Author – R. Sakthivel, Department of Statistics, Presidency College, Chennai, India
Second Author – Martin L. William, Department of Statistics, Loyola College, Chennai, India